DOCUMENT RESUME

ED 187 179 HE 012 571

AUTHOR Ramagli, Howard J., Jr.; Greenwood, Gordon E.
TITLE The Doctor Fox Effect: A Paired Lecture Comparison of Lecturer Expressiveness and Lecture Content.
PUB DATE 80
NOTE 34p.; Paper presented at the Annual Meeting of the American Educational Research Association (Boston, MA, April 1980). Best copy available.

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Affective Behavior; Classroom Research; *College Faculty; *College Instruction; College Students; Comparative Analysis; Higher Education; *Lecture Method; Self Expression; *Student Evaluation of Teacher Performance; *Teacher Behavior; Teacher Effectiveness
IDENTIFIERS Goffman (Erving)

ABSTRACT

The influence of the Doctor Fox effect on student ratings on instruction was examined. The idea for the Doctor Fox effect stemmed from the work of Erving Goffman and his notion that expressive behavior may influence an audience as much or more than substance when there is little time or reason for the audience to evaluate the presentation (1950). Undergraduate subjects compared two lecture segments varying in content and expressiveness. Total rating scores for each paired lecture were analyzed separately, along with a lecture choice variable and an adjective checklist. Results indicate a strong influence on student ratings of instruction by expressive teaching behaviors. Student choice of better lecture presentation and of adjectives to describe the lecturer showed strong relationships with high expressiveness lectures. The validity of the use of student ratings as the sole measure for administrative decisions regarding college teaching is questioned. (SW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED187179

The Doctor Fox Effect:
A Paired Lecture Comparison of Lecturer Expressiveness and Lecture Content

Howard J. Ramagli, Jr.

and

Gordon E. Greenwood

University of Florida

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Gordon E. Greenwood

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

HE 012 501

American Educational Research Association

Boston, 1980

Abstract

The influence of the Doctor Fox Effect on student ratings on instruction was examined. Undergraduate subjects compared two lecture segments varying in content and expressiveness. Total rating scores for each paired lecture were analyzed separately, along with a lecture choice variable and an adjective checklist. Results indicated a strong influence on student ratings of instruction by expressive teaching behaviors. Student choice of better lecture presentation and of adjectives to describe lecturer showed strong relationships with high expressiveness lectures. The study questions the validity of the use of student ratings as the sole measure for administrative decisions regarding college teaching.

Another Assessment of the Dr. Fox Effect: Paired Lecture Comparisons of Expressiveness and Content

The increased use of student ratings of instructors has caused concern in many sectors of higher education. Student rating systems are now the most common form of evaluation of college teachers. These ratings are attempts to hold college teachers accountable for their teaching effectiveness.

The Use of Student Ratings

One issue raised in relation to student ratings of instruction is their continued, increased use in administrative decision-making, specifically regarding faculty promotion, tenure, and salary increases.

Seldin (1975) surveyed colleges and universities across the country as to whether or not they used student ratings and secondly, how they used them, if they used them. Over one-half of the liberal arts colleges stated that they used student ratings, in some form or another, to evaluate faculty performance. Further, over 60% of all colleges and universities stated that the primary use for their student ratings was to evaluate faculty teaching performance for use in administrative decision-making.

Boyd and Schietinger (1976) surveyed colleges and universities in the southern region of the U.S. for the Southern Regional Education Board. Their results confirmed the trends cited in Seldin (1975) and showed that, at least in Southern colleges and universities, the trends were becoming more wide spread. They found that 62% of doctoral institutions used student ratings of instruction for administrative decision-making with regard to the faculty.

In addition, they found that large universities had little or no concern for the use of student ratings of instruction for faculty development programs. Still further, they found that the primary users of student ratings for decision-making purposes were academic deans and department chairmen. Finally, they found that 88% of all the institutions surveyed used student ratings of instruction for decision-making.

This use of student ratings for decision-making has prompted comments from many individuals and researchers. For example, McKeachie (1969) raised the question of the use of student ratings for personnel decision-making and contended that students were qualified, perhaps the most qualified, to evaluate teaching effectiveness. Centra (1973) reviewed the problem of the use of student ratings for decision-making with regard to retention, promotion, and tenure. He reported a survey in which students stated that they believed that their ratings of instruction should be the basis, at least in part, for decisions relating to promotion. Greenwood and Renner (1975) reviewed the literature on the use of student ratings and point out that the question of the validity of student ratings places their use in question. Further, they call attention to the fact that there is some agreement that some kind of student outcome measure, i.e. student achievement data in some form, should be the basis of validation for student ratings of instruction. Yet, they concluded that the ambivalence of the results of validation research raises questions of the reasonableness for the use of student ratings in administrative decision-making.

The Doctor Fox Literature

The Doctor Fox effect originally appeared in the literature in 1970 (Ware et al., 1970 a and b). The idea for the Doctor Fox effect stemmed from the work of Erving Goffman; specifically his notion that expressive

behavior may influence an audience as much or more than substance when there is little time or reason for the audience to evaluate the presentation (Goffman, 1959). There are two basic premises behind the Doctor Fox effect. One premise is that highly expressive behavior of a teacher will have primary influence on and be a determining factor of student ratings of instruction. The second premise is an outgrowth of the first premise. It states that students will be unable to distinguish between lectures of different levels of content when exposed to a lecturer with a highly expressive lecturing style.

The first of the Doctor Fox studies was reported by Naftulin et al. (1973). Their results stated that three groups of experienced educators were satisfied with the amount they had learned in a lecture presented by Dr. Fox, despite the fact that the lecture itself was low in content coverage (and high in expressiveness). Kaplan (1974) criticized this original study on two fronts:

(1) that it was not possible to determine if the Doctor Fox effect was due to the style of the lecture presentation or due to the impressive credentials given to Dr. Fox, and

(2) that there was no item on the lecture evaluation questionnaire which asked about learning, that is to say that there was no way to establish whether the subjects had the illusion of having learned.

The Doctor Fox studies are an extension of the above research and criticism; these studies took an experimental approach to the problem of the effect of the expressiveness of teachers. They did this by varying the amount of expressiveness with amount of content coverage in a lecture.

Ware and Williams (1975a) reported that student ratings of an instructor

were not sensitive to the amount of information covered in lectures, even though student achievement was affected. Specifically they found that, under high seductive (expressiveness) conditions, the students could not distinguish between a high content coverage lecture and a low content coverage lecture. Further they found that difference in lecture content coverage and lecturer seductiveness appeared to affect both student ratings and student test performance. In another reporting of this effect, Ware and Williams (1975b and 1977) found that differences in faculty information-giving are easier to detect using student rating scores when faculty members are less enthusiastic (low in seduction) in their lecture presentations. Also, they reported a statistical confirmation (using discriminant function analysis) of all the above results, that is the majority of error in detecting differences in lecture content was a result of bias in student ratings due to seductiveness of presentation.

Williams and Ware reported two efforts in attempting to "see through" the Doctor Fox effect. In the first (1976), they provided students with varying amounts of monetary incentive for learning in the lectures. They concluded that sensitizing students to the content of the lectures did not improve student accuracy in the rating of content coverage. In the second case (1977), they tried exposing students to two consecutive lectures of the same conditions, i.e. providing students with additional exposure to the lecturer does not enhance the students' ability to "see through" the Doctor Fox effect.

Throughout all the above studies, one result has remained constant. Lecturer expressiveness has a major impact on student ratings of instruction, and in every study reported, there was a significant difference in the rating

of high expressiveness lectures as opposed to low expressiveness lectures. This result is the primary premise of the Doctor Fox effect.

Two recent studies have confirmed this primary premise (Perry et al., 1979; Meier and Feldhusen, 1979). However, Frey (1979) has questioned these results by contending that bias in student ratings, due to this primary premise, can be reduced by altering student ratings instruments so that they contain more items geared toward learning outcomes. Ware and Williams (1979) have responded to this criticism by pointing out that Frey's criticism, based on his focus that specific items relating to content coverage be included in Doctor Fox research, is not valid. They pointed out specific items (knew subject matter well, increased student knowledge of the subject, inspired confidence in knowledge of the subject, and organized and presented the subject matter well) on their rating instrument which show Frey's criticism to be in error. Further, they contended that their results and the results of Meier and Feldhusen (1979) showed that students are unable to distinguish content level under varying experimental conditions. Finally, they reaffirmed their original contention that the Doctor Fox effect dealt primarily with the influence of expressiveness on student ratings, and that content recognition by students was a secondary issue.

The purpose of this study was to examine the following research questions: (1) Do levels of lecturer expressiveness and levels of lecture content interact with each other in affecting student evaluation of lecturer effectiveness? (2) After viewing a second lecture presentation varying in content and/or expressiveness from the first lecture presentation, do the levels of content and levels of expressiveness of this second lecture interact with the levels of content and levels of expressiveness of the first lecture

presentation in affecting the student evaluation of lecturer effectiveness? (3) Are there differences in student evaluation of lecturer effectiveness for lectures which are high in expressiveness as opposed to those which are low in expressiveness? (4) Does the level of content in a lecture presentation affect student evaluation of lecturer effectiveness? (5) Are there differences in student choices of adjectives to discribe what they felt was a better lecture presentation?

Method

Subjects

The 278 subjects for this study were obtained from those students registered in undergraduate courses of the Foundations of Education Department at the University of Florida.

The median age of the subjects was 19.9 years. Of the 278 subjects, 83.8 percent were females and 16.2 percent were males; their median grade point average was the equivalent of a B+. Juniors comprised 54.7 percent of the sample; seniors, 25.2 percent of the sample; sophomores, 17.3 percent of the sample; and postgraduates and freshmen, 2.8 percent of the sample.

Each of the sixteen classes was randomly assigned to one of the sixteen lecture pairs stated above. Thus, this is a random assignment of intact classes. Individual assignment for randomization was not done and certain precautions were taken before conducting the analyses of the data.

These precautions were to insure that the partitioning of the between groups variance is not confounded by the classes nested within treatments, nor by the unequal sizes of the classes. The second consideration is dealt with by conducting a chi-square analysis on the class size frequencies as they were assigned to the treatment groups. A significant chi-square would mean that there was a difference in the class sizes which would confound the

interpretation of results.

This test was made and a x^2 of 7.566 was obtained. This value with 9 degrees of freedom is not significant. Thus the problem of confounding of group size does not need to be considered in reporting the results.

The fact that classes are nested within the treatment groups was also considered. If the variance attributable to these nested classes is large enough to be significant when tested against the variance for the subjects within the groups, the interpretation of results will be confounded. This was tested by conducting a 2 x 2 analysis of variance with the classes nested for Lecture A. An F value of .78 at 12 and 262 degrees of freedom was obtained. This value is not significant and the classes nested within groups does not need to be partitioned out of the total variance for the model tested in the Lecture A analysis.

The Lectures

The lectures used in this study were the original Dr. Fox lectures as used by Ware (1974). These lectures are on videotape and are copies of the original lectures. However, instead of using the entire twenty minute videotaped lecture as did Ware, only the first five minutes of a given lecture were used. To prevent viewer fatigue, and yet to allow enough viewing time for comparison, the same five munute segment from each of the lectures was used.

Each lecture segment used had a specific level of content and a specific level of expressiveness; both content and expressiveness could be either high or low. The combinations of content levels and expressiveness levels produced four lecture segments. These were, including their average length of time in parentheses, as follows:

High content - High expresiveness (5 minutes, 6.45 seconds)

High content - Low expressiveness (5 minutes, 28.29 seconds)

Low content - High expressiveness (5 minutes, 3.68 seconds)

Low content - Low expressiveness (5 minutes, 25.66 seconds)

These lecture segments were then paired together in all possible combinations, including each one of the four lectures followed by itself, and were edited together on the videotape. Each lecture segment in the pair was preceded on the videotape by a fifteen second segment which displayed a sign on the screen identifying the lecture segment as either "Lecture A" or "Lecture B". The total number of combinations of these four lecture segments is sixteen. These combinations also take into account that the order of presentation of the lecture segments must be controlled for, and thus each pair of lecture segments has its reversed order pair present in the sixteen pair combinations.

The Content Dimension

The content dimension is defined by Ware (1974) as containing a specified number of teaching points for the conditions of the content dimension. A high content lecture contained (if the entire lecture were used) 26 teaching points. The low content lecture contained (if the entire lecture were used) 4 teaching points. In this study the low content lecture segment used 0 teaching points.

The Expressiveness Dimension

Expressiveness has been defined by Ware (1974) as follows:

> ...a number of different terms have been used, all of which appear to refer to the seduction-expressiveness aspect of teaching effectiveness, including: "charisma", "dynamism", "emotional appeal", "friendliness", "interest", "personality", "seduction", "stimulation", and "warmth". Seduction-expressiveness is operationally defined as the extent to which these mannerisms are present in the lecture presentation. (p. 4)

Thus, the high expressive lecture contains all or most of the elements of the above definition of expressiveness. The low expressive lecture contains few, if any, of the above elements.

In addition to the above information on the expressiveness dimension, Ware (1974, pp. 61-64) gives a detailed description of the training of the actor to display the expressiveness elements cited above.

Dependent Variables

The dependent measures for this study were a total student rating score (the same rating instrument as in the original Doctor Fox studies), a lecture choice variable, and a choice/no choice on each of the adjectives on an adjective checklist.

The student rating instrument was the same instrument used in the original Doctor Fox studies (Ware, 1974; Ware & Williams, 1975). It is a modified student evaluation instrument as developed by Pohlman (1975), with an internal consistency of total score rating of .96, using the KR-20 formula. This rating instrument was adapted by the original Doctor Fox researchers from a pre-existing instrument and is designed to measure the characteristics of the expressiveness and content dimensions of the Doctor Fox effect.

The reliability analysis for the student rating instrument for this study is a pooled KR-20 reliability across treatment groups. This pooling was done to avoid the between group differences, caused by any treatment effect, which might affect the reliability of the instrument.

The pooled reliability for the student rating instrument on the Lecture A presentation was .9425. The pooled reliability for the student rating instrument on the Lecture B presentation was .9530.

The lecture choice instrument is a simple, dichotomous forced choice question. The students were asked to select which of the two lecture

presentations they saw was the better presentation. They were asked to check a box corresponding to Lecture A or a box corresponding to Lecture B. They were asked to choose only one of the two lectures. Of the 278 subjects, 277 responded as directed to this question.

The lecturer adjective checklist consisted of forty-seven adjectives. These adjectives were drawn from (the literature) on characteristics of effective teachers (Bridges, et al., 1971; Hildebrand et al., 1971; and Wotruba and Wright, 1975).

The characteristics identified were compiled into one list and all multiple word descriptions were reducted to single word adjectives. This list was then compared to the descriptions in Ware (1974, pp. 4, 16-18, and 62-63) of the expressive or seductive lecturer. All adjectives which were not in the Ware sources were rejected as inappropriate. The remaining adjectives were then given a random order and the checklist was complete.

Procedure

The following procedure and administrative directions were used by the researcher for each of the sixteen classes used in this study:

> Good morning. My name is Howard Ramagli. This morning I would like to ask you to help me. In an effort to develop new ways of providing learning experiences for students, we are field testing some videotaped lectures as possible supplementary material for introductory classes in the area of psychology. The specific lectures we are testing are on the biochemistry of memory. What we are asking you to do today is to help us evaluate the quality of lecturing in these videotaped lectures. You will be shown two segments of tapings of a lecture on the biochemistry of memory. After you view each lecture segment, you will be asked to evaluate

the quality of lecturing in that particular segment. Let me repeat this. I will show you a segment of Lecture A and then I will ask you to evaluate the quality of lecturing in that segment. I will then show you a segment of Lecture B and you will be asked to evaluate the quality of lecturing in that particular segment. Are there any questions? Please open to page 1 of your booklet. The evaluations we are collecting from you will be kept confidential. We do need some descriptive information on you as a group. This information will also be kept confidential. Would you please fill in the information requested on page 1. When you have finished, please do not turn the page. Please look up at me to indicate you have finished.

.....

(Subjects completed descriptive information)

Is everyone finished?

.....

Now that all of you are finished, we will begin. We will start with a segment from Lecture A. (TURN ON VIDEOTAPE) This is Lecture A.

.....

(Subjects viewed Lecture A)

Please turn to page 2 in your booklet. This is the evaluation form we would like for you to fill out for Lecture A. (READ INSTRUCTIONS) (Discuss item 1 and 8, explaining their five choices to circle) Please respond to all of these items. Please begin. When you have finished, please do not turn the page. Please look up at me to indicate you have finished.

.....

(Subjects complete rating instrument for Lecture A)

Is everyone finished?

.....

Now that you all are finished, we will continue. We will now see a segment from Lecture B. (TURN ON VIDEOTAPE) This is Lecture B.

.....

(Subjects viewed Lecture B)

Please turn to page 4 in your booklet. This is the evaluation form for Lecture B. It is the same form you filled out for Lecture A. Please complete this form. Please respond to all of these items. When you have finished, please do not turn the page. Please look up at me to indicate you have finished.

.....

(Subjects complete rating instrument for Lecture B)

Is everyone finished?

.....

Now that you have finished, please turn to page 6. The instructions ask you to select which of the two lecture segments you saw was the better lecture presentation. Please place a check in the box below either Lecture A or Lecture B. Please do not check them both. When you have finished, please look up at me.

.....

(Subjects complete lecture choice variable question)

Please turn to page 7. On this page you will find a list of adjectives. Please read the instructions carefully. They ask you to select those adjectives, characteristic of the instructor, which were influential in your choice of which lecture presentation was better. Please place a check on the line to the left of each adjective, characteristic of the lecturer, which was influential in your choice of the lecture segment you chose. You may select as many or as few adjectives as you choose. When you have finished, please look up at me.

.....

(Subjects complete adjective checklist)

Now that you have finished, please pass your booklets forward. I would like to thank each of you for your help. We will be conducting this evaluation with a number of classes in the College of Education. In an effort to get the honest opinion of all of those individuals we will be asking to evaluate the videotapes, we ask you please, do not discuss what you have done or seen with anyone outside of this class. This is most important if we are to get an honest evaluation of the quality of these videotapes. I will be glad to come back, if you are interested, toward the end of the quarter and provide you with feedback as to the results of these evaluations. Thank you for helping me.

Design

A 2 x 2 (Content x Expressiveness) factorial design was used to answer research questions relating to Lecture A. A 2 x 2 x 2 x 2 (Content, Lecture A x Expressiveness, Lecture A x Content, Lecture B x Expressiveness, Lecture B) factorial design was used to answer questions relating to Lecture B. Because the number of subjects per cell was not equal and this unequal n was the result of natural groups of unequal size, an unweighted means analysis of variance solution is necessary to analyze the data (Timm and Carlson, 1975). All data were analyzed using the appropriate SAS procedure. A Chi Square analysis was used to answer research questions relating to the Lecture Choice variable and each adjective on the adjective checklist.

Results

Results for the Lecture A analyses yield a significant two-way interaction Results for the Lecture B analyses yield one significant two-way interaction but no three-way or four-way interactions. Means and standard deviations for the two analyses are reported in tables 1 and 2, 5 and 6 respectively.

A summary of the respective ANOVA's can be found in Table 7. The adjective checklist results are found in Table 8.

Lecture A Analyses

These analyses yielded the following results: (1) a significant interaction between content and expressiveness, (2) a significant difference toward higher ratings of the high content versus the low content lectures when expressiveness was high, and (3) a significant difference toward higher ratings of higher expressiveness versus low expressiveness lectures for both levels of content.

Lecture B Analyses

The analyses of Lecture B yielded the following results: (1) A significant interaction between content and expressiveness for Lecture B, (2) a significant difference toward higher ratings of high expressiveness versus low expressiveness lectures for both levels of content for Lecture B, and (3) the high and low expressiveness preconditions for Lecture A produced a significant difference in the ratings of Lecture B for both precondition content levels.

Lecture Choice Variable

The chi-square analysis of the lecture choice variable yielded the following results: (1) a significant difference in the frequency of choice of Lecture B over Lecture A and (2) a trend toward the selection of high expressiveness lectures over low expressiveness lectures.

Adjective Checklist

The chi-square analysis of the adjectives on the adjective checklist yielded the following results: (1) 21 of the 47 adjectives had significant difference in choice versus no choice and (2) 18 adjectives (dramatic, energetic, inspiring, expressive, forceful, interesting, stimulating, interested

warm, motivating, enthusiatic, showmanship, enjoyable, humorous, sparkling, dynamic, witty, and persuasive) showed a strong relationship with the high expressiveness lectures.

Conclusions

Overall, the results of this study seem to point in one direction, confirmation of the primary premise of the Doctor Fox Effect. In other words, the results showing differences on the student ratings of the lectures and on the lecture choice variable due to the expressiveness dimension for both levels of content confirms the Doctor Fox Effect research previously mentioned. Further, it appears that the expressiveness dimension does have a primary influence on the student ratings of instruction.

The results of this study seem to indicate that previous experience, or the opportunity to compare lecturing styles, also may influence evaluation of the teaching process by students. A careful examination must be made of the result of the significant difference between the ratings of high and low expressiveness lectures produced by the expressiveness precondition of the first lecture. This result showed that low expressiveness of the first lecture influenced higher student ratings of the second lecture and that high expressiveness of the first lecture influenced lower student ratings of the second lecture. This result was true for both precondition content levels. Since there were no significant interactions between the preconditions of the first lecture and the treatment conditions of the second lecture, it can be concluded that the expressiveness precondition of the first lecture presentation has an independent influence on the student ratings of the second lecture presentation. This precondition expressiveness influence is independent of the influence that the content and expressiveness levels of the second lecture presentation have on the student ratings of that second lecture presentation.

The results of Ware and Williams (1975a) and Williams and Ware (1977) indicated a secondary premise, or corollary, to the Doctor Fox effect. The conclusion, drawn from their results, was that under a condition of high expressiveness, lecture content would have no effect on student ratings; under low expressiveness conditions, lecture content would have an effect on student ratings. The results of this study do not support entirely this conclusion. In fact, for the analysis of the first lecture, the exact opposite of this conclusion was found. For the first lecture, the results showed no difference in student ratings for content at low expressiveness. However, at high expressiveness, there was a significant difference in the rating of the high content versus the low content lectures. However, this significance difference disappeared for the analysis of the second lecture.

The explanation of this reversal of the content dimension results may lie in the shortness of the lecture presentation. It may be that five minutes is not enough time to ascertain fully the content present in the lecture segment. Rather, the rating difference may be an artifact of the content by expressiveness interaction. This explanation is supported by the fact that the result disappears when the second lecture results are examined. Thus, this shortness of time may focus the attention of the students on the more easily recognizable quality in lecturing style; that of expressiveness.

This focusing on the expressiveness of the lecture by students is supported additionally by the lecture choice and adjective checklist results. Students had a tendency to select high expressiveness over low expressiveness lectures. Also, 18 out of the 21 adjectives with significant frequency distributions showed a tendency to be related to high expressiveness lectures. These same 18 adjectives were descriptive of the expressive lecturer. In fact, 15 of the 18 adjectives came directly from Ware's (1974) definition of

expressiveness. These factors seem to be, at least, a confirmation that students can identify high expressive behaviors when selecting adjectives which influence their choice of a best or better lecture presentation. The conclusion from the facts above is that students' selection of which lecture was better is based primarily on high expressive behaviors.

Discussion

The implications of this study relate to two issues. Those issues are the increased use of student ratings of instruction by college administrators for personnel decision-making and the validity of student ratings of instruction. Since the use of student ratings hinges on the validity question, the issue of the validity of student ratings is considered first.

The results of this study imply that the primary influence on student ratings of instruction is expressive lecturing style and behaviors. In addition, this study implies that students, when comparing lecture presentations varying in content and expressiveness are not able to identify differences in content level, or at the very least, become unaware of content level in a lecture presentation. These implications raise serious questions about the validity of student ratings. Supposedly student ratings provide an assessment of the student's perception of a faculty member's ability to teach content. Yet, the conclusions of this study are that student perception is influenced primarily by, not what the professor presents, but how he presents it. Thus, student ratings are providing an assessment only of student perception of teaching style, and this is only part of the picture.

The validity of student ratings can not be assessed completely if student reporting of faculty teaching is biased because of expressive teaching style. Student ratings do not provide all of the information necessary for administrators to assess faculty teaching performance, let alone for

administrators to make personnel decisions about faculty promotion, tenure, and salary increases.

The increased use of student ratings is questionable in the light of the above implications from the results of this study. If it is in fact the case that faculty performance is being judged primarily on the basis of student ratings, then faculty members with a non-expressive teaching style are being punished in cases where student ratings are used primarily for personnel decisions. Further, students can be fooled or seduced into believing a faculty member is a good teacher solely by his exhibition of a highly expressive teaching style, regardless of the amount of information that the faculty member may be imparting to students. Thus, it may be possible to reward a faculty member, who teaches his students little or no information, because he has a highly expressive teaching style and high student ratings as a result of that teaching style.

References

Boyd, J. E. and Schietinger, E. F. Faculty evaluation procedures in southern colleges and universities. Atlanta: Southern Regional Education Board, 1976.

Bridges, C. M., Ware, W. B., Brown, B. B., and Greenwood, G. E. Characteristics of best and worst college teachers. Science Education, 1971, 55(4), 545-553.

Centra, J. A. The student as godfather: The impact of student ratings on academia. Educational Researcher, 1973, 2(10), 4-8.

Frey, P. W. The Doctor Fox effect and its implications, Instructional Evaluation, 1979, 3(2), 1-5.

Greenwood, G. E. and Renner, R. R. Student ratings of college teaching: Their validity and use in administrative decision-making. Science Education, 1975, 59(4), 493-498.

Goffman, E. The presentation of self in everyday life. New York: Doubleday Books, 1959.

Hildebrand, M., Wilson, R., and Dienst, E. Evaluating university teaching. Berkeley: Center for Research and Development in Higher Education, University of California, 1971.

Kaplan, R. M. Reflections on the Doctor Fox paradigm. Journal of Medical Education, 1974, 49, 310-312.

McKeachie, W. J. Student ratings of faculty, AAUP Bulletin, 1969, 55, 439-443.

Meier, R. S. and Feldhusen, J. F. Another look at Doctor Fox; Effect of stated purpose of evaluation, lecturer expressiveness, and density of lecture content on student ratings. Journal of Educational Psychology, (in press).

Naftulin, D. H., Ware, J. E., and Donnelly, F. A. The Doctor Fox Lecture: A paradigm of educational seduction. Journal of Medical Education, 1973, 48, 630-635.

Perry, R. P., Abrami, P. C., and Leventhal, L. Educational seduction: The effect of instructor expressiveness and lecture content on student ratins and achievement. Journal of Educational Psychology, 1979, 71, 107-116.

Pohlman, J. T. A multivariate analysis of selected class characteristics and student ratings of instruction. Multivariate Behavioral Research, 1975, 10, 81-92.

Timm, N. H. and Carlson, J. E. Analysis of variance through full ranks. Multivariate Behavioral Research Monographs, 1975, MBR Monograph No. 75 - 1.

Ware, J. E. The Doctor Fox effect: A study of lecturer effectiveness and ratings of instruction. Doctoral dissertation, Southern Illinois University, Carbondale, 1974.

Ware, J. E., Donnelly, F. A., and Naftulin, D. H. The evaluation of medical education. Journal of Medical Education, 1970, 45(10), 813. (a)

Ware, J. E., Donnelly, F. A., and Naftulin, D. H. The evaluation of medical education. In The complete papers on the ninth annual conference on research in medical education. Washington, D.C.: The Association of American Medical Colleges, 1970, 131-135. (b)

Ware, J. E. and Williams, R. G. The Doctor Fox effect: A study of lecturer effectiveness and ratings of instruction. Journal of Medical Education, 1975, 50, 149-156. (a)

Ware, J. E. and Williams, R. G. Studies of student-faculty rating scoring methods: Seeing through the Doctor Fox effect. Fourteenth Annual Conference on Research in Medical Education during the eighty-sixth Annual Meeting of the AAMC, Washington, D.C., 1975. (b)

Ware, J. E. and Williams, R. G. Discriminant function analysis of student rating as a means of identifying lecturers who differ in enthusiasm or information giving. Education and Psychological Measurement, 1977, 37, 627-639.

Ware, J. E. and Williams, R. G. Seeing through the Doctor Fox effect: A response to Frey. Instructional Evaluation, 1979, 3(2), 6-10.

Williams, R. G. and Ware, J. E. Validity of student ratings of instruction under different incentive conditions: A further study of the Doctor Fox effect. Journal of Educational Psychology, 1976, 68(1), 48-56.

Williams, R. G. and Ware, J. E. An extended visit with Dr. Fox: Validity of student ratings of instruction after repeated exposures to a lecturer. American Educational Research Journal, 1977, 14(4), 449-457.

Wotruba, T. R. and Wright, P. L. How to develop a rating instrument: A research proposal. Journal of Higher Education, 1975, 46(6), 653-663.

Table 1

Summary of Means & Standard Deviations for Lecture A

Group	N	Mean	Standard Deviation
High Content			
High Expressiveness	68	56.5735	14.2466
Low Expressiveness	90	41.1333	11.0293
Low Content			
High Expressiveness	62	68.8710	14.1160
Low Expressiveness	58	41.5172	12.2564

Table 2

Summary of Means & Standard Deviations for Lecture B

Group	N	Mean	Standard Deviation
High Content (Lecture B)			
High Expressiveness (Lecture B)			
High Content (Lecture A)			
High Expressiveness (Lecture A)	17	60.1765	12.3806
Low Expressiveness (Lecture A)	17	71.5294	9.9443
Low Content (Lecture A)			
High Expressiveness (Lecture A)	16	66.4375	11.1593
Low Expressiveness (Lecture A)	11	64.1818	14.4279
Low Expressiveness (Lecture B)			
High Content (Lecture A)			
High Expressiveness (Lecture A)	17	40.9412	13.9260
Low Expressiveness (Lecture A)	28	41.6786	11.6684
Low Content (Lecture A)			
High Expressiveness (Lecture A)	17	44.1675	12.9819
Low Expressiveness (Lecture A)	20	49.9500	14.7522
Low Content (Lecture B)			
High Expressiveness (Lecture B)			
High Content (Lecture A)			
High Expressiveness (Lecture A)	22	57.7727	15.0931
Low Expressiveness (Lecture A)	29	64.5172	14.0804
Low Content (Lecture A)			
High Expressiveness (Lecture A)	12	53.0000	12.9263
Low Expressiveness (Lecture A)	14	66.7143	12.0667

Table 2 (contd)

Low Expressiveness (Lecture B)			
High Content (Lecture A)			
High Expressiveness (Lecture A)	12	49.0000	11.5444
Low Expressiveness (Lecture A)	16	51.2500	12.7619
Low Content (Lecture A)			
High Expressiveness (Lecture A)	17	40.0588	11.4098
Low Expressiveness (Lecture A)	13	49.0769	14.3786

Table 3

2 x 2 Analysis of Variance for Student Rating Instrument Total Score on Lecture A

ANOVA for 2 x 2 Design Source	F
Between Groups (Model)	22.80*
Content	5.49**
Expressiveness	53.28**
Content x Expressiveness	6.71**
Simple Main Effects	
Expressiveness of High Content	56.05**
Expressiveness at Low Content	9.73**
Content at High Expressiveness	11.69**
Content at Low Expressiveness	.03

*($p < .05$, $F_{(3,274,.05)} = 2.64$)

**(p .05, $F_{(1,274,.05)} = 3.88$)

*** (< .01, $F_{(1,274,.01)} = 6.74$)

Table 4

2 x 2 x 2 x 2 Analysis of Variance for Student Rating Instrument Total Score on Lecture B

Source	F
Between Groups (Model)	11.25*
Content (Lecture A)	.06
Expressiveness (Lecture A)	13.43*
Content (Lecture B)	.35
Expressiveness (Lecture B)	114.44*
Content (A) x Expressiveness (A)	.16
Content (A) x Content (B)	3.48
Content (A) x Expressiveness (B)	.10
Expressiveness (A) x Content (B)	1.56
Expressiveness (A) x Expressiveness (B)	.83
Content (B) x Expressiveness (B)	6.51
Content (A) x Expressiveness (A) x Content (B)	2.98
Content (A) x Expressiveness (A) x Expressiveness (B)	2.04
Expressiveness (A) x Content (B) x Expressiveness (B)	.26
Content (A) x Content (B) x Expressiveness (B)	2.68
Content (A) x Expressiveness (A) x Content (B) x Expressiveness (B)	2.13

Table 4 (contd)

Content at High Expressiveness	5.21
Content at Low Expressiveness	2.02
Expressiveness at High Content	95.07**
Expressiveness at Low Content	33.98**

*p < .05

**p < .01

Table 5

Student Rating Instrument Mean Scores on Lecture B Collapsed into Lecture B Treatment Groups

Group	N	Mean
High Content		
High Expressiveness	61	65.5813
Low Expressiveness	77	44.1843
Low Content		
High Expressiveness	82	60.5011
Low Expressiveness	58	47.3464

Table 6

Student Rating Instrument Mean Scores on Lecture B Collapsed into Lecture A Precondition Treatment Groups

Group	N	Mean
High Content		
High Expressiveness	68	51.9726
Low Expressiveness	90	57.2438
Low Content		
High Expressiveness	62	50.9159
Low Expressiveness	58	57.4807

Table 7

Sample Population Frequency Distribution on Lecture Choice

Lecture Choice	Lecture Type* HC - HE	HC - LE	LC - HE	LC - LE
Lecture A	34	10	23	6
freq. %	12.3	3.6	8.3	2.2
Lecture B	58	44	69	33
freq. %	20.9	15.9	24.9	11.9
$x^2_{(3)}$ =	9.542	$p < .05$		

*HC = High Content
LC = Low Content
HE = High Expressiveness
LE = Low Expressiveness

LC - x^2 = 9.542
$p < .05$ (.0229) 3df

HC - x^2 = 3.048
3df
n.s.

Table 8

Chi-Square Values for Adjectives on Lecture Choice

Adjective	Variable Chi-Square	Adjective	Variable Chi-Square
Dramatic	47.378*	Enthusiastic	96.341*
Knowledgeable	14.774*	Showmanship	12.017*
Thorough	4.521	Vital	4.894
Coherent	5.477	Concise	5.680
Well-prepared	4.026	Lucid	1.058
Energetic	94.109*	Enjoyable	20.780*
Logical	13.378*	Clear	1.900
Explanatory	7.286	Inventive	3.977
Inspiring	16.973*	Cultured	7.773
Understandable	4.164	Confident	6.118
Well-paced	3.993	Poised	4.398
Expressive	48.246*	Humorous	49.172*
Forceful	31.729*	Sparkling	14.145*
Interesting	22.463*	Competent	5.134
Variety	1.986	Charismatic	8.122
Stimulating	27.030*	Conscientious	1.089
Interested	13.093*	Dynamic	26.206*
Warm	14.862*	Well-groomed	0.180
Scholarly	4.404	Organized	1.935
Orderly	19.220*	Careful	6.671
Stylized	5.910	Witty	24.496*
Motivating	27.598*	Unique	5.627
Masterful	1.192	Persuasive	11.354*

*p < .01, df = 3